

Simulation of stochastic network dynamics via entropic matching

Tiago Ramalho,^{1,*} Marco Selig,² Ulrich Gerland,¹ and Torsten A. Enßlin²

¹*Arnold Sommerfeld Center for Theoretical Physics*

(ASC) and Center for Nanoscience (CeNS),

LMU München, Theresienstraße 37, 80333 München, Germany

²*Max Planck Institute for Astrophysics (MPA),*

Karl-Schwarzschild-Straße 1, 85741 Garching, Germany

Abstract

The simulation of complex stochastic network dynamics arising, for instance, from models of coupled biomolecular processes remains computationally challenging. Often, the necessity to scan a models' dynamics over a large parameter space renders full-fledged stochastic simulations impractical, motivating approximation schemes. Here we propose an approximation scheme which improves upon the standard linear noise approximation while retaining similar computational complexity. The underlying idea is to minimize, at each time step, the Kullback-Leibler divergence between the true time evolved probability distribution and a Gaussian approximation (entropic matching). This condition leads to ordinary differential equations for the mean and the covariance matrix of the Gaussian. For cases of weak nonlinearity, the method is more accurate than the linear method when both are compared to stochastic simulations.

*Electronic address: tiago.ramalho@physik.uni-muenchen.de

I. INTRODUCTION

Stochastic dynamical systems arise in many areas of science, with much recent interest focusing on biomolecular reaction networks that control the behavior of individual cells [1–4]. For instance, the expression levels of different genes in a single cell are often coupled by direct or indirect regulation and are subject to intrinsic (reaction) noise as well as extrinsic (parametric) noise [5]. Models of such systems can be based on a wide range of different approaches and mathematical tools [6, 7]. A full treatment of noise effects on the single-molecule level requires an approach based on the chemical Master equation [8]. However, many noise phenomena can also be understood and quantitatively analyzed within a more coarse-grained continuum description based on stochastic differential equations or, equivalently, the associated Fokker-Planck equations [7–9]. For instance, such a description is adequate to describe the excitable system behavior of a natural genetic circuit whose excitations are triggered by noise [10], or the synchronizing effect of a coupling between noisy synthetic “repressilators” [11, 12].

A recurrent problem in the analysis of these systems is that the quantitative behavior of the theoretical models must be computed not only once, but for a large number of different parameter values. One common situation is that, for a given model, one would like to determine a “phase diagram” of system behaviors over the entire parameter space, in order to understand the model on a theoretical level. Similarly, when models are leveraged for the interpretation of experiments, model fitting to data is required to infer system parameters or to discriminate between different variants of the model. In quantitative biology, there is currently a clear need for systematic and efficient methods to reverse engineer models from a limited number of noisy experimental observations [13, 14]. Any such method requires as an essential ingredient an efficient technique for the forward simulation of the observables. For discrete stochastic models based on the chemical Master equation, recent progress in this direction has been made by using spectral methods [15]. Also, efficient sampling techniques have been developed for rare event problems [16, 17]. Here, we focus on continuum models and develop an approximation scheme that can efficiently capture the stochastic dynamics of the “typical events” even in larger dynamical systems.

For a given nonlinear stochastic system, we consider the dynamics of its probability distribution over state space. Our scheme is reminiscent of the linear noise approximation

[8, 9] in that it approximates the probability distribution at each point in time by a Gaussian distribution. The essential difference is in the criterion used to match the Gaussian approximation to the true distribution. In our method, we take a more global stance at the matching process, by invoking the Kullback-Leibler divergence [18] as accuracy measure. We minimize the Kullback-Leibler divergence at each point in time to obtain a set of ordinary differential equations for the parameters of the Gaussian distribution which best approximates, in this sense, the current probabilistic state of the system. We refer to our matching condition as entropic matching. It was recently proposed as a general method to construct simulation schemes for partial differential equations [19]. Here, we develop and test this idea for the case of stochastic differential and Fokker-Planck equations. For the examples considered here, we find the method to be more accurate and robust than the linear noise approximation, while the computational complexity is comparable. In the following, we first formulate the theoretical problem, then describe the method in detail, and finally illustrate and test it on simple example systems.

II. PROBLEM STATEMENT

We consider a nonlinear stochastic system with N variables c_i with $i = 1, \dots, N$. We think of the c_i as concentration variables for different molecular species in a biomolecular network, however the method developed below is not specific to these systems. The interactions between the molecular species are specified by a set of dynamical equations, which depend on a set of parameters, which we collectively denote as a . Throughout this article, whenever we omit an index, we refer to the vector containing all elements of the respective set. The dynamical equations can be written as stochastic differential equations (SDE)

$$\dot{c}(t) = \frac{d}{dt} c(t) = f(c(t), a) + \eta(t) , \quad (1)$$

where f is the nonlinear function regulating each node of the biomolecular network and η represents a random process characterizing the system's intrinsic fluctuations. We can think of a single trajectory obtained by solving Eq. (1) (with one realization of the noise process $\eta(t)$) as the biochemical trajectory of a single cell, and an ensemble of such trajectories as the corresponding dynamics for a population of cells. This picture relies on the assumption that different cells' trajectories are uncorrelated, i.e. that there is no intercellular communication.

Within the framework of SDE's, extrinsic cell-to-cell noise would need to be accounted for by specifying a probability distribution for the parameters a , which can either be static or have a dynamics by itself. For simplicity, we will ignore extrinsic noise in the following. From the SDE system (1) one can then directly obtain [8, 9] a Fokker-Plank Equation (FPE) of the form

$$\partial_t P(c, t) = -\partial_c [f(c, a) P(c, t)] + \frac{1}{2} \partial_c^2 [\mathcal{X} P(c, t)] , \quad (2)$$

where \mathcal{X} represents the covariance of the random process denoted by η in Eq. (1). Instead of deriving the FPE from a phenomenological set of SDE's, one can also start from the chemical Master equation for a given set of microscopic biomolecular reactions and then approximate the discrete stochastic process by a continuous drift and diffusion process [8, 9].

Either way, given a FPE for our biomolecular network, one central task is to be able to determine the values for the parameters a for which the stochastic model best describes experimental data that is taken at a set of time points t_j . For instance, within a Bayesian approach to infer the parameter values from the data, it is necessary to calculate the posterior density

$$P(a|c, t) = \frac{P(a) \prod_j^J P(c|a, t_j)}{\int da P(a) \prod_j^J P(c|a, t_j)} \quad (3)$$

where $P(a)$ represents the prior knowledge about the parameters of the system. Thus, the posterior density can be obtained once we have calculated the “forward solution” $P(c|a, t_j)$ for all relevant time points. The direct solution of this problem is to simulate uncorrelated paths of the system given by Eq. (1) using an appropriate discretization procedure [20] and reconstruct the probability density function for the whole population. This approach is very computationally heavy and thus unsuitable for large systems. However, in our test applications below (which are small systems), we use this exact approach as a reference.

The proposed method relies on approximating the true distribution for $P(c|a, t_j)$ with a Gaussian distribution

$$P(c|a, t_j) = \mathcal{G}(c - \bar{c}(t), C(t)), \quad (4)$$

$$\mathcal{G}(c - \bar{c}, C) = \frac{1}{\sqrt{|2\pi C|}} e^{-\frac{1}{2} (c - \bar{c})^\dagger C^{-1} (c - \bar{c})}. \quad (5)$$

Here, $|C| = \det(C)$ and $c^\dagger d = \sum_{i=1}^N c_i d_i$ is a scalar product over the space of network state vectors. In the Gaussian approximation, knowledge of the temporal dynamics of the

dynamics of its average \bar{c} and covariance matrix C is sufficient, which simplifies not only analytic but also numerical calculations. The proposed method thus consists of calculating $\bar{c}(t)$ and $C(t)$ such that the Gaussian approximation best represents $P(c|a, t)$ in an information theoretic sense. The resulting equations can then be efficiently computed numerically and the resulting approximation used to infer the desired parameters. How well the true non-Gaussian probability distribution can be approximated by distribution (4) is critical to the performance of our algorithm, as we will see.

III. METHOD

A. Background

Since we are approximating $P(c|a, t_j)$ by the Gaussian (4), only the first two moments need to be calculated for all times. Hence we seek closed ordinary differential equations (ODEs) for the first two moments, without requiring the calculation of the full probability distribution. They are given by two functions, $f_{\bar{c}}$ and f_C such that:

$$\begin{aligned}\dot{\bar{c}}(t) &= f_{\bar{c}}(\bar{c}(t), C(t)) \\ \dot{C}(t) &= f_C(\bar{c}(t), C(t))\end{aligned}\tag{6}$$

The usual approach is to linearize the function f in Eq. (2). Then, assuming the initial condition is itself Gaussian, the evolution of the probability distribution consists of nothing more than a series of linear transformations. Thus it is possible to quickly compute linearized evolution equations for the Gaussian approximation:

$$\begin{aligned}\dot{\bar{c}}(t) &= f(\bar{c}) \\ \dot{C}(t) &= J(t)C(t) + C(t)J^T(t) + \mathcal{X}(t)\end{aligned}\tag{7}$$

where J is the Jacobian of the function f at the current concentration $c(t)$. In our numerical examples below, we use this linear noise approximation as a reference for comparison to our proposed method.

With our method, we now aim to derive evolution equations of the form (6) without an *a priori* linearization of f , which will better preserve the characteristics of the original nonlinear system.

B. Proposed method

We will construct $f_{\bar{c}}$ and f_C using the concept of entropic matching of evolving distributions [19]. A general outline of the method is to assume an initial Gaussian distribution for c , evolve it according to the nonlinear function f and then find the Gaussian distribution which best matches it in an information theoretic sense. This allows for the derivation of ODEs for the Gaussian parameters conforming to the aforementioned criterion.

An initial Gaussian distribution $P(c) = \mathcal{G}(c - \bar{c}, C)$ at time t characterized by $\bar{c} = \bar{c}(t)$ and $C = C(t)$ will evolve according to the dynamics described by the nonlinear function f . For now we assume this evolution to be deterministic. Intrinsic noise shall be dealt with later. After an infinitesimal time step δt , the evolved distribution function $P'(c')$ will be:

$$P'(c') = \mathcal{G}(c - \bar{c}, C)|_{c=c' - \delta t \dot{c}} \left| \frac{dc}{dc'} \right|, \quad (8)$$

according to the rule of change of variables for probabilities, where $c(t + \delta t) = c' = c + \delta t \dot{c}$ and thus $|dc'/dc|_{c=c'} = |1 + \delta t df/dc|_{c=c'}$ to linear order in δt . $P'(c')$ is in general not of Gaussian functional form. Obviously a representation such as $P'(c') = \mathcal{G}(c' - \bar{c}', C')$ with an updated $\bar{c}' = \bar{c}(t')$ and $C' = C(t')$ at $t' = t + \delta t$ is no longer exact. However, it is possible to obtain values for \bar{c}' and C' such that the information content of $P'(c')$ is represented as well as possible. Information theoretical considerations [21–23] single out the Maximum Entropy Principle for this. This principle is equivalent to requiring a minimal Kullback-Leibler divergence [18] of $\mathcal{G}(c' - \bar{c}', C')$ to P' , or a minimal relative Gibbs free energy [24]. All these measures (relative entropy, Kullback-Leibler divergence, and Gibbs free energy) can be regarded as a measure of the information theoretical distance between two distributions, although they are not a metric in the strict mathematical sense due to their asymmetry with respect to the different roles of the matching and matched distributions.

To obtain such a maximally informative Gaussian we entropically match the parameters of the Gaussian \mathcal{G}' to the time evolved distribution P' by minimizing the relative entropy (or Kullback-Leibler divergence):

$$S(\mathcal{G}', P') = - \int dc' \mathcal{G}(c' - \bar{c}', C') \log \frac{\mathcal{G}(c' - \bar{c}', C')}{P'(c')} \quad (9)$$

We can consider this expression as a functional to be minimized with respect to all degrees of freedom of our matching distribution, or simply as a function of the two parameters c' and

C' to be minimized with respect to their finite number of degrees of freedom. The minimum of this functional will define a Gaussian distribution which has maximal information content about the time evolved distribution P' subject to a deterministic nonlinear evolution.

Defining $S(\mathcal{G}', P') \equiv S(\bar{c}', C')$ since the Kullback-Leibler divergence is now a function only of the Gaussian parameters, Eq. (9) is explicitly

$$\begin{aligned}
S(\bar{c}', C') &= \\
&- \int dc' \mathcal{G}(c' - \bar{c}', C') \log \frac{\mathcal{G}(c'_i - \bar{c}', C')}{\mathcal{G}(c_i - \bar{c}(t), C(t))|_{c_i=c'_i - \delta t \dot{c}_i} \left| \frac{dc_i}{dc'_i} \right|} \\
&= \int dc' \mathcal{G}(c' - \bar{c}', C') \left[\frac{1}{2} \log \frac{|C'|}{|C|} + \log \left| 1 - \delta t \frac{df}{dc} \right|_{c=c'} \right. \\
&\quad \left. + \frac{1}{2} (c' - \bar{c}')^T C'^{-1} (c' - \bar{c}') - \frac{1}{2} (c - \bar{c})^T C^{-1} (c - \bar{c})|_{c=c' - \delta t \dot{c}} \right]
\end{aligned} \tag{10}$$

where P' was expressed as a function of $c = c' - \delta t \dot{c}$. By discarding terms of order δt^2 , the entropic divergence between the Gaussian \mathcal{G}' and the time evolved former Gaussian P' , Eq. (9), can be brought into the form

$$\begin{aligned}
S(\bar{c}', C') &\simeq \frac{1}{2} \log \left(\frac{|C|}{|C'|} \right) + \frac{1}{2} \text{tr} [C' C^{-1} - \mathbb{1} \\
&\quad + \delta t (2 - C^{-1} C' - C' C^{-1}) \left\langle \frac{df}{dc} \right\rangle_{c'}] \\
&\quad + \frac{1}{2} (\bar{c}' - \bar{c})^T C^{-1} (\bar{c}' - \bar{c}) - \delta t (\bar{c}' - \bar{c})^T C^{-1} \langle f(c') \rangle_{c'},
\end{aligned} \tag{11}$$

where $\langle \cdot \rangle_{c'}$ denotes the expectation value over $\mathcal{G}'(c') = \mathcal{G}(c' - \bar{c}', C')$. The desired values of \bar{c}' and C' are the minima of Eq. (11).

To take intrinsic noise into account, we model it as a Gaussian of zero mean and δt accumulated covariance matrix. This assumption is not necessary but is general enough for our purposes and allows us to go further with the calculations analytically. We define this noise as:

$$\mathcal{X}_{kl} \delta t = \langle \xi'_k \xi'_l \rangle = \int_0^{\delta t} dt' dt'' \langle \xi_l(t') \xi_k(t'') \rangle \tag{12}$$

To take these fluctuations into account we consider the final probability distribution to be a sum of this random variable with the maximally informative Gaussian, which corresponds to a convolution:

$$P(c') = \int d\xi \mathcal{G}(c' - \bar{c}' + \xi, C') \mathcal{G}(\xi, \mathcal{X} \delta t) \quad (13)$$

$$= \mathcal{G}(c' - \bar{c}', C' + \mathcal{X} \delta t), \quad (14)$$

where \bar{c}' and C' were obtained by the entropic matching process. This noise addition in an Itô-like fashion provides the final Gaussian approximation for time $t' = t + \delta t$ including intrinsic noise. This procedure can then be applied again and again to generate subsequent time steps, creating an iterative scheme to track the evolution of $\bar{c}(t)$ and $C(t)$. Having such a scheme, we can take the limit $\delta t \rightarrow 0$ and obtain an ODE for the time evolution of \bar{c} and C of the form of Eq. (6), from which $f_{\bar{c}}$ and f_C can be read off.

Minimizing Eq. (11) with respect to $\bar{c}'(t)$ and $C'(t)$ we obtain simple equations for the evolution of both averages and covariances. By adding the required noise term to the evolution of the covariance matrix and taking the limit $\delta t \rightarrow 0$ we obtain the final ODEs for the evolution of the Gaussian approximation of our system:

$$\begin{aligned} \dot{\bar{c}} &\simeq \langle f(c) \rangle \\ \dot{C} &\simeq \left\langle \frac{df}{dc} \right\rangle_c C + C \left\langle \frac{df}{dc} \right\rangle_c^T + \mathcal{X} \end{aligned} \quad (15)$$

These are the general equations we sought. Since the Gaussian expectation value of a nonlinear function is usually analytically inaccessible in a closed form we now develop an approximation method which will make numerical calculations tractable.

One may expand the nonlinear function determining the dynamics around its average as:

$$f(c) = \sum_{n=0}^N \frac{1}{n!} \left. \frac{d^n f}{dc^n} \right|_{c=\bar{c}} (c - \bar{c})^n \quad (16)$$

$$\langle f \rangle_c = f(\bar{c}) + \frac{1}{2} \left. \frac{d^2 f}{dc^2} \right|_{c=\bar{c}} C + \dots \quad (17)$$

$$\left\langle \frac{df}{dc} \right\rangle_c = \left. \frac{df}{dc} \right|_{c=\bar{c}} + \frac{1}{2} \left. \frac{d^3 f}{dc^3} \right|_{c=\bar{c}} C + \dots \quad (18)$$

This is justified when the spread of the PDF around its average is small and highlights why in many cases the linear approximation works, as it corresponds to the first order terms in this series and is thus valid in regimes where noise is very small. In index notation we can write the expansion as:

$$\langle f_i \rangle_c = f_i(\bar{c}) + \frac{1}{2} \frac{d^2 f_i}{dc_k dc_l} \Big|_{c=\bar{c}} C_{kl} + \dots \quad (19)$$

$$\left\langle \frac{df_i}{dc_j} \right\rangle_c = \frac{df_i}{dc_j} \Big|_{c=\bar{c}} + \frac{1}{2} \frac{d^3 f_i}{dc_j dc_k dc_l} \Big|_{c=\bar{c}} C_{kl} + \dots \quad (20)$$

If we were to take only the first order in the expansion (16) we would recover the commonly used linear noise approximation, Eq. (7). Our derivation however shows that the correct equation for linearised evolution includes the full Gaussian expectation values and thus we expect a significant improvement in performance by taking additional terms in this series. In a practical case it appears that with only the second order approximation significant improvements are already visible over the linear approximation, because the correlation structure is taken into consideration in the evolution of the parameters. In our numerical examples presented in the following section, we use the expansion up to second order as in Eq. (19).

Thus, when the method is applicable, it will be several orders of magnitude faster than a stochastic simulation. It remains to show under which situations it is applicable. In the next section we will investigate the performance of the algorithm with some numerical examples.

IV. TEST APPLICATIONS

The ODEs of the entropic matching algorithm were integrated with the Adams-Moulton method. The usage of this implicit integration scheme is appropriate here due to its numerical robustness, which can successfully deal with the stiffness of the ODEs in the general nonlinear case. The results were compared to stochastic simulations of the system dynamics using the stochastic analogue of the Euler scheme and to the linear noise approximation (7).

A. Van der Pol oscillator

The van der Pol oscillator was chosen as an initial test case. Here it is possible to exactly calculate $\langle f(c) \rangle_c = \int dc f(c) \mathcal{G}(c - \bar{c}, C)$, the expectation value of the nonlinear function f describing deterministic part of the dynamics due to the vanishing of all derivatives in its Taylor expansion of order three and higher. Thus, the only approximation made in this case

is the assumption of Gaussianity. Furthermore, a single parameter controls the nonlinearity strength, allowing one to investigate the algorithm's performance for different regimes.

The van der Pol dynamics are described by:

$$\ddot{x}_i = \mu(1 - x_i^2)\dot{x}_i - \omega_i^2 x_i + \sum_{j \neq i} \gamma_{ij}(x_j - x_i) + \xi_i, \quad (21)$$

where μ controls the nonlinearity, γ determines the connectivity of the network and ω controls individual oscillation frequency of each node. $\xi \leftarrow \mathcal{G}(\xi, \mathcal{X})$ represents white noise intrinsic to the system (in the numerical experiments only diagonal covariance matrices were considered for the white noise). By doing the usual substitution $y = \dot{x}$ it is possible to bring the dynamics into a form suitable for numerical integration. It then becomes clear that the fourth order derivative tensor vanishes everywhere and thus it is possible to calculate the expectation value of the nonlinear function exactly with little computational effort by taking only the first two terms of the expansion; cf. Eq. (17). The parameter values used for the simulation were $\gamma_{i,j} = 0$ except $\gamma_{0,1} = 2$, $\gamma_{1,0} = 5$, $\gamma_{2,1} = 3$; $x_0 = 2, 0, 0$; $y_0 = 0, 1, 2$; $\omega = 2, 1, 2$. The intrinsic noise (12) was characterized by a diagonal covariance matrix with diagonal entries set to 0.1.

As a term of comparison we simulated the system for a quasi-linear case, $\mu = 0.05$, where all three methods should perform identically. This is clear from the second panel of Fig. 1, where the averages coincide perfectly for the visualized species. All three methods used the same initial values for the averages. The covariance matrix was initialized as the value of the intrinsic noise level for both our method and the linear noise method, while in the stochastic simulation all paths begin at the same initial value and therefore the initial variance is null. In the third panel, where the standard deviation is plotted for one of the species, it is clear that after some time all three methods also produce identical values.

The stochastic simulation consisted of 10^3 runs of the system with time step $\delta t = 10^{-3}$. Such a small time step was necessary to ensure correct results with the Euler method. The other methods consisted of a single run of the Adams-Moulton method with $\delta t = 10^{-3}$. In this case, however, decreasing δt to 10^{-1} had no effect in the accuracy of the obtained trajectories. It must be noted that for the Gaussian methods the elements of the covariance matrix must be followed in addition to the average, meaning that we must follow $\frac{3N+N^2}{2}$ numbers instead of N . In practice, the Gaussian methods were two orders of magnitude

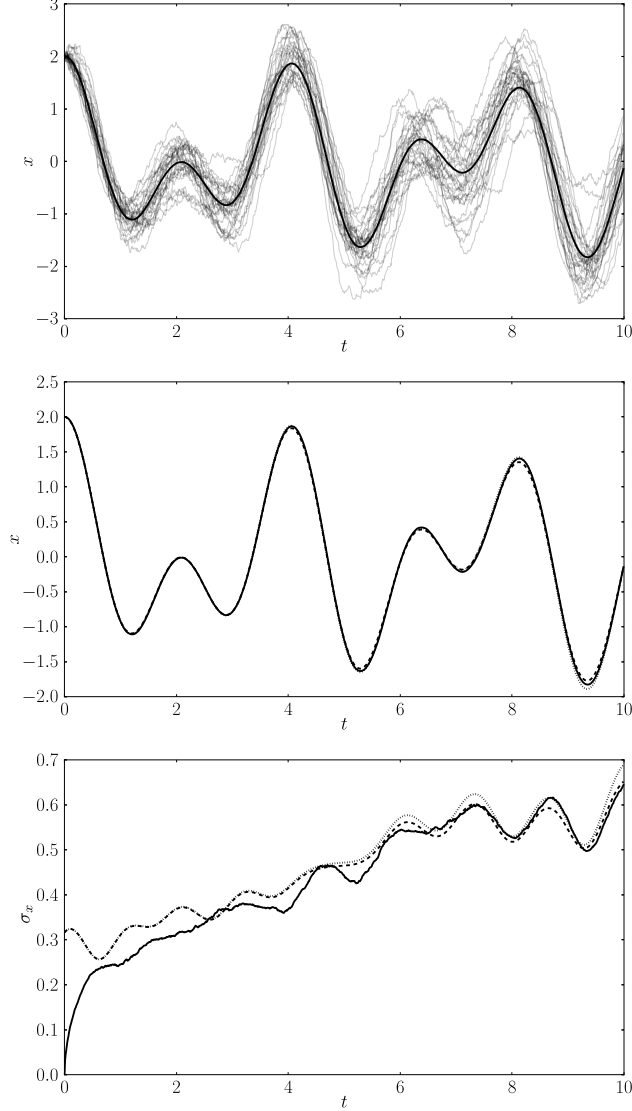


FIG. 1: First panel, stochastic simulation with 1000 runs (some individual runs are plotted) of system (21) and corresponding average value for the first species of a system with $\mu = 0.05$ and $N = 3$. Second panel, comparison of stochastic average value (solid) with prediction using Eq. (15) (dashed) and Eq. (7) (dotted). Third panel, comparison of standard deviation for the stochastic simulation (solid) with prediction using Eq. (15) (dashed) and Eq. (7) (dotted).

faster than the stochastic simulations at the same time step (which could be safely reduced for the former, yielding an order of magnitude further speed increase).

In the first three panels of Fig. 2 it is clear all algorithms perform well since the true probability distribution remains approximately gaussian even at a late time $t = 10$. Since

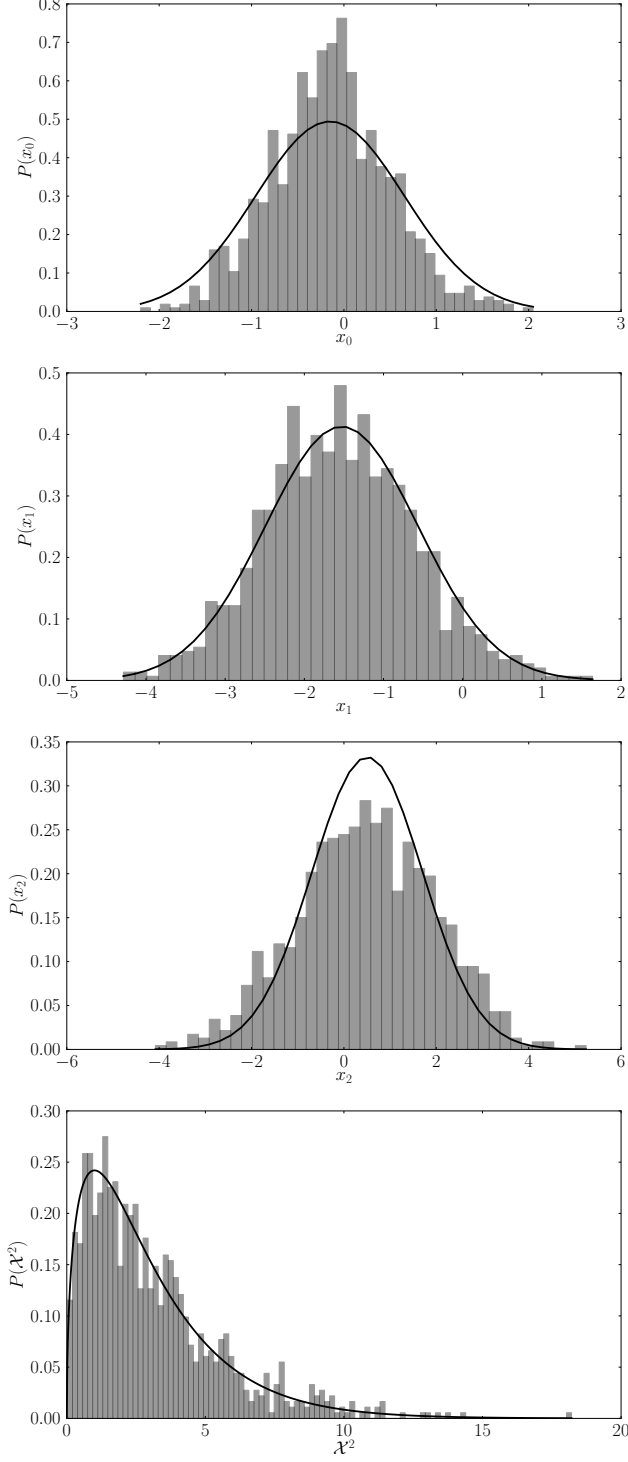


FIG. 2: Final histogram of stochastic simulation with 1000 runs of system (21) with $\mu = 0.05$ and $N = 3$ and corresponding prediction using Eq. (7) for time $t = T$, in this case $T = 10$. First three panels represent the three components of the system. Last panel, histogram of chi squared distribution for the stochastic simulation data compared to the χ^2 distribution with 3 degrees of freedom.

only projections of multivariate gaussians can be visualized, we calculated the χ^2 probability distribution, which can help us verify if the full distribution (i.e., taking into account the full correlation matrix) does indeed match a Gaussian with the predicted parameters. If the prediction matches the simulation, a histogram of $\chi^2 = (Y - \bar{c}_p)^\dagger C_p^{-1} (Y - \bar{c}_p)$, where \bar{c}_p and C_p are the predicted average and covariance, should match the chi squared distribution for a Gaussian $\mathcal{G}(\bar{c}, C)$ with N degrees of freedom, where N is the number of dimensions of the system. In the fourth panel of Fig. 2 it is clear that this is the case for $\mu = 0.05$.

The case with $\mu = 1.5$ is analyzed next. The system now exhibits clear nonlinear behavior, although it is not predominant. Comparing the results of the entropically matched scheme with those of the linear noise approximation, Fig. 3, the average trajectory approximates the stochastic simulations better. Dampening due to the progressive loss of synchronization in individual stochastic trajectories is apparent only in our method. This is due to the nonlinearity and is evident by looking at the second order term in Eq. (17) where the Hessian contains terms proportional to $-\mu x_i$ and $-\mu y_i$ and is multiplied by the covariance matrix. Thus there is a damping proportional to the magnitude of the fluctuations.

Regarding the covariance matrix (third panel in Fig. 3), in the linear noise approximation the estimate for the variance oscillates around the true value, under or overestimating it, until finally losing synchrony. In contrast, our method consistently underestimates the variance (inherent to the Gaussian approximation, since the true distribution exhibits "fat tails", Fig. 4, first panel) but always remains synchronous with the true value. Thus we consider the estimate produced by entropic matching to be more robust. It has been verified that in some cases the covariance matrix diverges in the linear noise approximation while the entropically matched scheme still produces reasonable results.

Regarding the histograms, Fig. 4, we note that the distribution remains unimodal for two components of the system and bimodal for one. Even though the unimodal distributions are non Gaussian our method still approximates their average and standard deviation well, which is the intended behavior. In the fourth panel it is visible that the multivariate gaussian no longer fully describes the system, due to the underestimation of the elements of the covariance matrix. However, the histogram still exhibits a quantitative shape which indicates the system is approximately Gaussian. If the target distribution would be completely non Gaussian, a flat histogram would be expected.

In the case of high nonlinearity ($\mu \gg 1$) the Gaussian approximation breaks down as

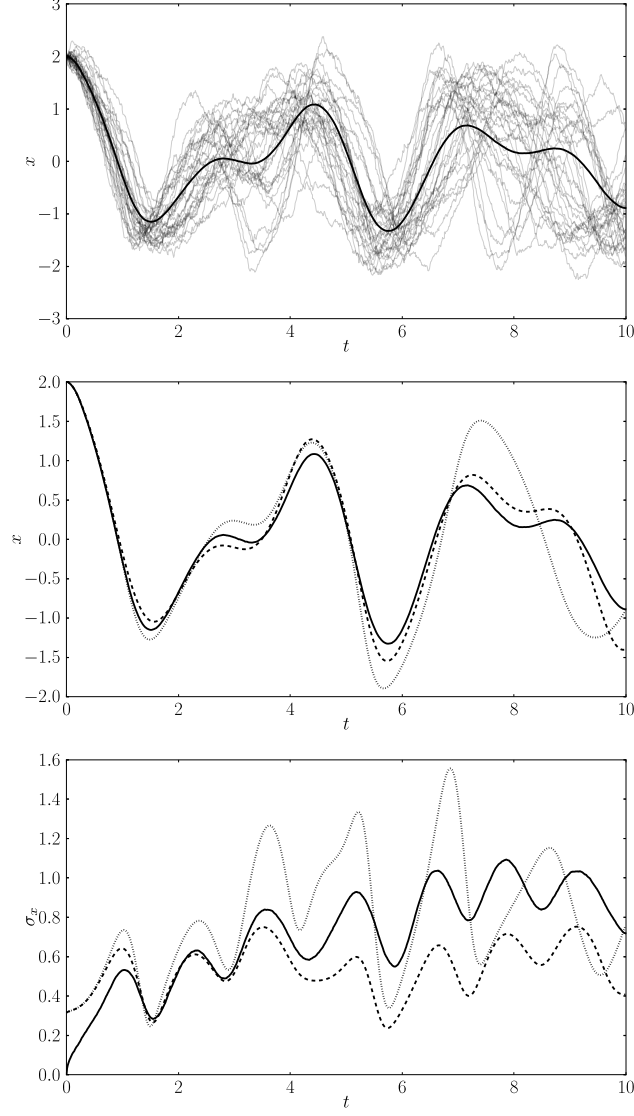


FIG. 3: First panel, stochastic simulation with 1000 runs (some individual runs are plotted) of system (21) and corresponding average value for the first species of a system with $\mu = 1.5$ and $N = 3$. Second panel, comparison of stochastic average value (solid) with prediction using Eq. (15) (dashed) and Eq. (7) (dotted). Third panel, comparison of standard deviation for the stochastic simulation (solid) with prediction using Eq. (15) (dashed) and Eq. (7) (dotted).

the tails of the target distribution gain significant probability mass and thus contribute more heavily to the trajectory, which means they cannot be ignored any more. Furthermore bimodal distributions arise, which cannot be accurately described by a multivariate Gaussian.

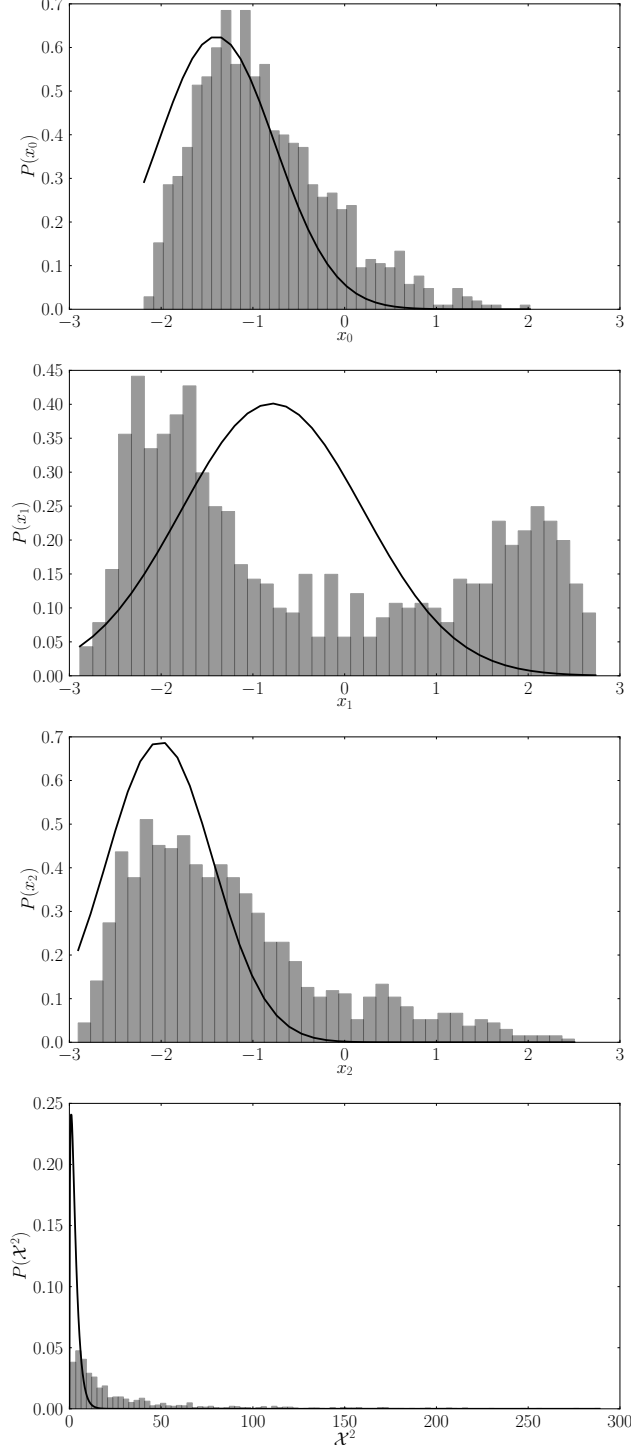


FIG. 4: Final histogram of stochastic simulation with 1000 runs of system (21) with $\mu = 1.5$ and $N = 3$ and corresponding prediction using Eq. (7) for time $t = T$, in this case $T = 10$. First three panels represent the three components of the system. Last panel, histogram of chi squared distribution for the stochastic simulation data compared to the χ^2 distribution with 3 degrees of freedom.

B. Gene network

As a final example, a simple genetic network was simulated. The simulated network consists of a system of mutually repressing genes inspired by the *repressilator* [11], which exhibits sustained oscillations for a wide range of parameter regimes. The dynamics of gene expression are simulated using hill type activation functions, where each gene is repressed by another in the circuit:

$$\dot{c}_i = \frac{k^n}{c_j^n + k^n} - \lambda c_i + \zeta_i \quad (22)$$

where $j = (i+1) \bmod N$, k is the half activation level, and n is the binding cooperativity (in a sense, the nonlinearity parameter).

However, in this case ζ_i is not Gaussian and c -independent noise but rather of Poissonian nature, which means that the noise covariance increases linearly with c . In this case it is still possible to apply the previously derived algorithm by defining a new variable $\rho = \log c$ such that

$$\dot{\rho}_i = \frac{k^n e^{-\rho_i}}{e^{\rho_j n} + k^n} - \lambda + \xi_i \quad (23)$$

where now $\xi \leftarrow \mathcal{G}(\xi, \mathcal{X})$. This means that ζ_i has a log normal distribution which has the property $\sigma_c^2 \propto \langle c \rangle$, which is more realistic.

The nonlinear function was Taylor expanded up to second order, which means that in this case there is also an approximation error in the expectation value of the function. However, it is sufficient to yield significantly better results than the linear noise approximation for low cooperativity (i.e., $n < 3$).

We have examined the case of $n = 2$ with uncorrelated intrinsic fluctuations of magnitude 0.1. As evidenced by Fig. 5, our method again outperforms a linear approximation at predicting the average value (middle panel). The standard deviation was overestimated, but as before, our estimate appears to be more robust. From the histograms in Fig. 6 we can see the distributions remain unimodal, which suggests that in this case the algorithm's performance could be improved by improving the series expansion used to represent the nonlinear function. However, as is evidenced by the χ^2 distribution in the last panel, the obtained estimate at the end of the simulation is relatively good.

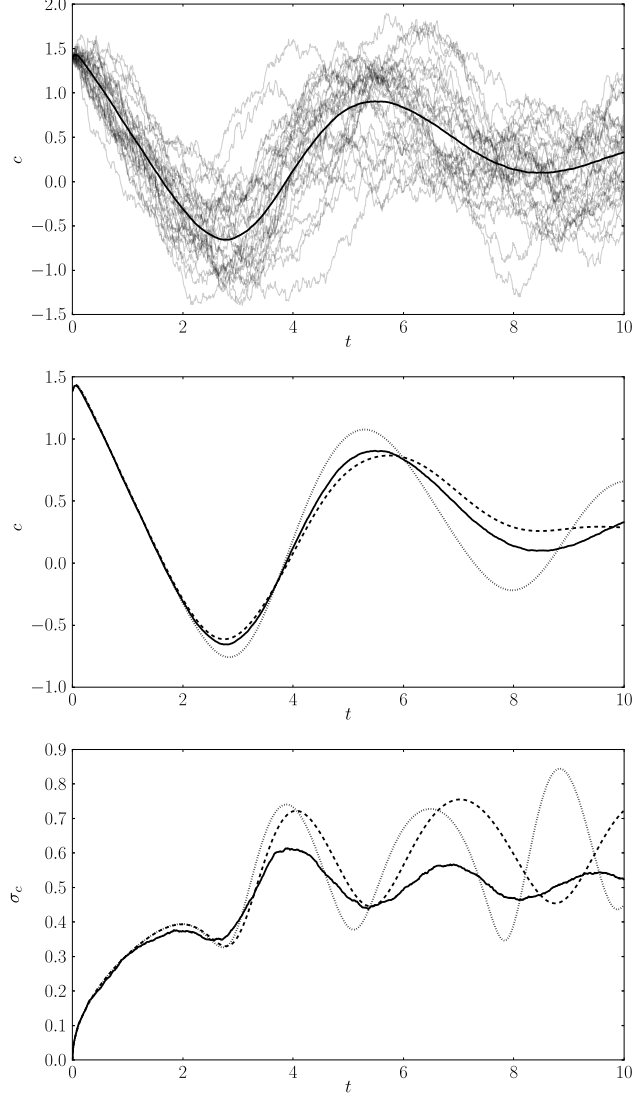


FIG. 5: First panel, stochastic simulation with 1000 runs (some individual runs are plotted) of system (23) and corresponding average value for the first species of a system with $n = 2$ and $N = 3$. Second panel, comparison of stochastic average value (solid) with prediction using Eq. (15) (dashed) and Eq. (7) (dotted). Third panel, comparison of standard deviation for the stochastic simulation (solid) with prediction using Eq. (15) (dashed) and Eq. (7) (dotted).

V. DISCUSSION AND CONCLUSIONS

This paper proposes an algorithm to predict the evolution of an ensemble of biochemical networks subject to stochastic fluctuations. The algorithm is based on entropic matching, a strategy developed in [19] to construct simulation schemes based on the information theo-

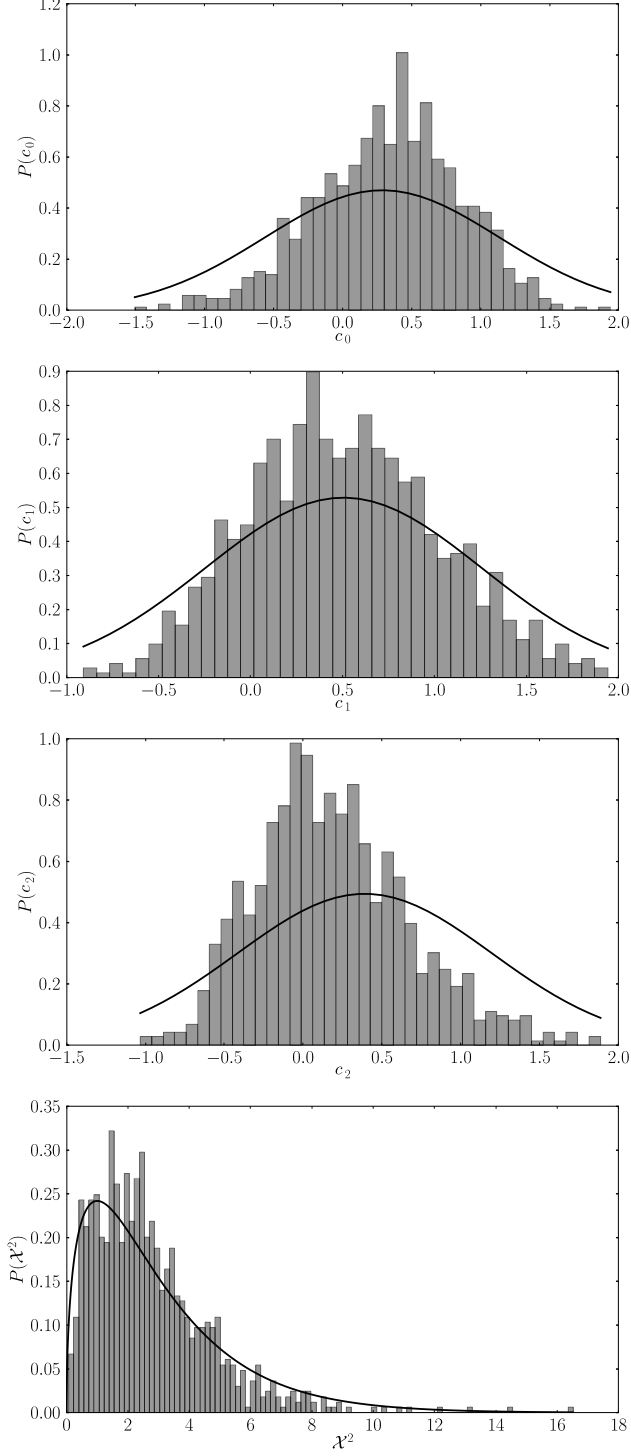


FIG. 6: Final histogram of stochastic simulation with 1000 runs of system (23) with $n = 2$ and $N = 3$ and corresponding prediction using Eq. (7) for time $t = T$, in this case $T = 10$. First three panels represent the three components of the system. Last panel, histogram of chi squared distribution for the stochastic simulation data compared to the χ^2 distribution with 3 degrees of freedom.

retical concept of the Maximum Entropy Principle. For each step in the time evolution of a system, the probability distribution describing the state of system is matched to a Gaussian and thus ODEs for the first and second moment of the distribution are derived (Eq.(15)). The evolution of these moments then tracks the average and covariance of the system parameters. We have implemented these equations numerically by resorting to a series method to compute the Gaussian expectation values (Eq. (19)).

By deriving the equations for the evolution of a Gaussian approximation to the full probability distribution without linearizing the system, a more accurate approximation method was developed which can be used in a wider parameter regime than the linear approximation with good results. In the cases where the system’s intrinsic noise is very low or the network dynamics are approximately linear, both methods work equally well. For moderate noise and weak nonlinearity, our method outperforms the linear approximation.

This algorithm proved to be much faster than a stochastic simulation due to a number of factors. To begin with, numerical algorithms for the solution of SDEs cannot guarantee convergence as fast as standard ODE routines. As an example, the commonly used stochastic Euler scheme has an order of convergence of merely 0.5 whereas the Adams-Moulton method we use below has an order of convergence of 4. Algorithms with even faster convergence orders are available and simple to implement. On the other hand, more advanced algorithms for SDEs yield marginally better convergence at the cost of high complexity in terms of implementation [20]. The upshot is that much smaller time steps must be used for a stochastic simulation, thereby increasing run time.

A second important point is the convergence of the stochastic method itself. To sample from the probability distribution of interest with good accuracy, one must obtain several thousands of paths from the aforementioned SDE simulation with variance scaling as $1/\sqrt{N_{\text{runs}}}$. Importance sampling might mitigate this problem somehow, but the number of paths required is still several orders of magnitude high. On the other hand, the currently proposed method requires only one evaluation.

A possible complication with the currently proposed method might be the calculation of higher order derivatives. However, current methods of algorithmic differentiation [25] allow a user to calculate derivative tensors of any order even when the derivatives are not analytically accessible. The calculation of these tensors is comparable in complexity to the calculation of the original function, being slower by only a fixed constant depending on the

order of the derivative.

In the case of strong nonlinearities, stochastic simulations are still the preferred method. Clearly, multimodal or highly asymmetric target distributions will be poorly approximated by a multivariate Gaussian. To remedy this, a possibility would be the use of a Gaussian mixture to represent the different modes of the probability distribution. However, this approach brings with it a number of problems as the entropic divergence cannot be computed analytically in this case. Additionally the number of parameters to be determined would scale at least as $L(N + \frac{N(N+1)}{2} + 1)$, where L is the number of Gaussian mixture components. This becomes unfeasible for a larger number of mixture components.

In practice, this procedure may be sufficient to derive an efficient inference scheme for stochastic nonlinear networks, an issue left for the future.

-
- [1] J. M. Raser and E. K. O’Shea, *Science* **309**, 2010 (2005).
 - [2] A. Eldar and M. B. Elowitz, *Nature* **467**, 167 (2010).
 - [3] G. Balázsi, A. van Oudenaarden, and J. J. Collins, *Cell* **144**, 910 (2011).
 - [4] B. Munsky, G. Neuert, and A. van Oudenaarden, *Science* **336**, 183 (2012).
 - [5] M. Elowitz, A. Levine, E. Siggia, and P. Swain, *Science* **297**, 1183 (2002).
 - [6] H. de Jong, *J. Comp. Biol.* **9**, 67 (2002).
 - [7] D. J. Wilkinson, *Nat. Rev. Genet.* **10**, 122 (2009).
 - [8] N. G. van Kampen, *Stochastic Processes In Physics And Chemistry* (Elsevier Science and Technology Books, 2007), 3rd ed.
 - [9] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer, 2010), ISBN 3642089623.
 - [10] G. M. Süel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz, *Nature* **440**, 545 (2006).
 - [11] M. B. Elowitz and S. Leibler, *Nature* **403**, 335 (2000).
 - [12] J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz, *Proc Natl Acad Sci USA* **101**, 10955 (2004).
 - [13] R. J. Prill, J. Saez-Rodriguez, L. G. Alexopoulos, P. K. Sorger, and G. Stolovitzky, *Science Signaling* **4**, mr7 (2011).
 - [14] B. Steiert, A. Raue, J. Timmer, and C. Kreutz, *PLoS ONE* **7**, e40052 (2012).

- [15] A. M. Walczak, A. Mugler, and C. H. Wiggins, Proceedings of the National Academy of Sciences **106**, 6529 (2009).
- [16] R. J. Allen, P. B. Warren, and P. R. Ten Wolde, Physical review letters **94**, 18104 (2005).
- [17] N. B. Becker, R. J. Allen, and P. R. ten Wolde, J Chem Phys **136**, 174118 (2012).
- [18] S. Kullback and R. Leibler, Annals of Mathematical Statistics **22** (**1**), 79 (1951).
- [19] T. A. Enßlin, ArXiv e-prints (2012), 1206.4229.
- [20] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, 1992), corrected ed., ISBN 3540540628.
- [21] E. T. Jaynes, Physical Review **106**, 620 (1957).
- [22] E. T. Jaynes, Physical Review **108**, 171 (1957).
- [23] E. T. Jaynes and G. L. Bretthorst, *Probability Theory* (Cambridge University Press, 2003).
- [24] T. A. Enßlin and C. Weig, Phys. Rev. E **82**, 051112 (2010), 1004.2868.
- [25] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation* (SIAM, 2008), ISBN 9780898716597.